Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 7, July 2021: 6698 - 6705

Research Article

Aspect Based Sentiment Analysis of Restaurant Reviews using Ensemble Algorithms

Shini George[1] and Dr.V. Srividhya[2]

1 Reasearch Scholar, Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu, India.

2 Assistant Professor, Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu, India.

shinijobi@gmail.com , vidhyavasu@gmail.com

Abstract:

Digital information is continuously generated from various sources such as social media, and user reviews for services. Analysing of these reviews to extract user opinions is critical for developing customer satisfaction. In particular, restaurant services may be improved with customer feedback if the user opinions or sentiments are inferred from user reviews. The development of automated software systems for evaluating customer reviews is ongoing. This research work proposes an ensemble machine learning approach on the basis of the lexicon method and machine learning classifiers and the performance of the new model is evaluated on the basis of accuracy. The Nave Bayes (NB), Decision Tree, and Support Vector Machine (SVM) are used to create an ensemble Classifier. The article presents a comparative assessment of the efficacy of ensemble methods for sentiment analysis. The restaurant review dataset is used to demonstrate the feasibility and benefits of the proposed approach. The work focuses on not only the overall opinions but also aspect-based opinions, including Food, Service, Ambience, Quality and Price. The analysis is followed by a wide range of comparative experiments demonstrating the efficacy of ensemble technology for sentiment analysis. Finally, some detailed discussion and conclusions are drawn concerning the effectiveness of ensemble technology.

Key Terms: Accuracy, Bagging, Stacking, Ensemble, Machine Learning, Support Vector machines.

1. INTRODUCTION

The Analysis of feelings, opinions, and attitudes about a subject or entity is called sentiment analysis (SA) or opinion mining (OM). The entity is covered by reviews. Extracting and analyses opinions of people about an entity is called Opinion Mining. While the feelings or emotions can be expressed in the form of the text then it will analyze which is also known as Sentimental Analysis.

Many customers or users use the review information for making the decision about the Movie, Restaurant, shopping website, etc. It is difficult for users to read and realize all the reviews about each restaurant on the website. In TripAdvisor website, there are many restaurants. Many users will look for which are the best in the website based on the reviews given by the customers. Each time there is a need to read all the reviews of all restaurants it will be time-consuming. Sentimental Analysis solve this problem by providing essential and necessary information.

This work analyses the data from restaurant reviews and applies different natural language processing techniques to reveal some important information that is apparently invisible to the viewer. With the help

of topic modelling techniques, customer reviews are classify based on some predefined aspects and it also help identify hidden topics over the raw data.

Several models use Machine learning classifiers for opinion mining. When observing this models it is clear that under some configuration they perform well and some other configuration the performance is worse. We propose an ensemble machine learning technique to solve this problem.

The ensemble learning technique creates a model by combining multiple classifiers. This method improve the classification accuracy [1]. The ensemble approach is a multimodal system that effectively combines several classifiers and techniques into a predictive model. Multiple models can be combined using Ensemble Methods to produce improved results.

There are six sections to this study. The second section provides background information about the selected topic and discusses related literature. In Section 3, we present the research methodology and related experiments to categorize restaurant reviews based on their aspect based sentiments. Section 4 focuses on experiments performance metrics. Section 5 discusses the experimental results. Finally, section 6 presents the summarized results and conclusion.

2. Related work

M. Govindarajan [2] propose homogeneous and heterogeneous hybrid approach with bagging and arcing respectively. SVM, Radial Basis Function, k-Nearest Neighbor and Multilayer perceptron are the base classifiers used in this work. Direct marketing dataset is used in this work. The homogeneous ensemble classifiers with bagging show better accuracy than single classifiers and heterogeneous ensemble classifiers provide higher accuracy than homogeneous ensemble classifier.

Araque et al., [3] first build a classifier using deep learning with the help of a linear machine learning algorithm and word embeddings model. This is conceded as a baseline for performance ovulation. This work proposes a set of ensemble methods and the performance of these model is evaluated with deep learning baseline. To evaluate this model, six public sentiment classifiers and six public datasets are used.

Yadav et al. [4] introduced a new hybrid classification system for the review of Restaurants based on coupling classification methods. This work analysed the effectiveness of the hybrid approach for opinion mining. The Classifiers Naïve Bayes (NB), Genetic Algorithm (GA) and Support Vector Machine (SVM) was used in this work. First, the NB, SVM, GA classifiers were constructed as the initial model, and lastly, an ensemble approach is suggested. The ensemble approach produces greater accuracy than the initial models. The test time of this model also decreases because of the data dimension reduction. The accuracy of Naïve Bayes was 85%, Support Vector Machine was 85.20%, Genetic Algorithm was 85.30% and the proposed Hybrid Method was 93%. This work proved that the ensemble method produces higher accuracy than single classifiers.

Sanjeev Kumar, and Dr. Ravendra Singh [5] performed a comparative analysis between single classifier and ensemble classifier. This work proved that ensemble classifiers outperformed the individual classifiers. Here J48 is used as the base classifiers and AdaboostM1, Bagging, and Stacking are the ensemble classifiers. The movie review data set is used for the analysis. Comparative analysis shows that the stacking performs better than other classifiers.

Azadeh Mohammadia and Anis Shaverizade [6] suggest deep neural network architecture based on ensemble techniques for aspect-based sentiment analysis. This method first build CNN, BiLSTM , GRU and LSTM deep learning models, after that stacking ensemble approach combines the outputs of these models. To combine the outputs of base classifiers logistic regression is used as a meta-learner. The proposed method increases 5 to 20% accuracy of aspect-based prediction.

3. Methodology

Several researchers have looked into combining multiple classifiers to create an ensemble classifier. The ensemble classifier increases robustness, accuracy, and overall generalization by combining redundant and complementary classifiers. This research analysis the usefulness of ensemble approaches for sentiment classification problems in depth. To predict classification scores, base classifiers such as Nave Bayes (NB), Decision tree Classifier, and Support Vector Machine (SVM) are built initially utilising VADER lexicon. We choice them because they are representative classification methods and in terms of strengths and philosophies of every heterogeneous technique. After that, employing base classifiers, the well-known homogeneous and heterogeneous ensemble technique is used to achieve excellent generalisation performance. The proposed methodologies' viability and benefits are proven using a restaurant review, which is widely applied in the field of sentiment categorization. A number of comparative studies have been carried out, and finally, an in-depth discussion and conclusions about the effectiveness of ensemble techniques for sentiment categorization are presented.

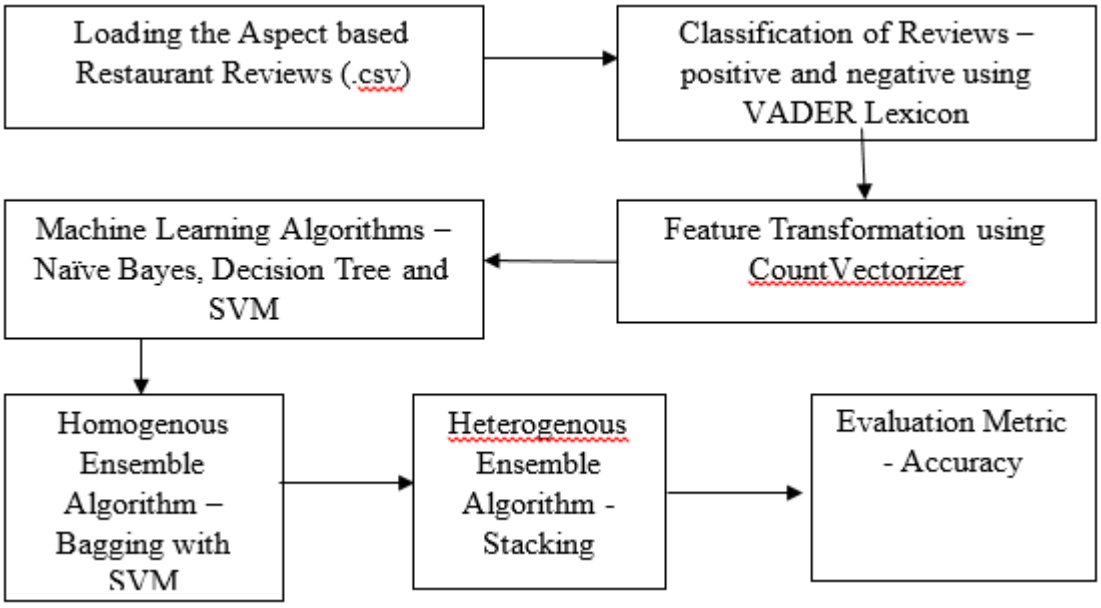

Figure 1 proposed methodology

This study suggests a hybrid strategy for solving the sentiment analysis challenge. The proposed strategy is focused on: Input the dataset, Lexicon based Classification, Feature Reduction phase, Single base classifiers, homogeneous classification and heterogeneous classification to generate the most accurate classification results. Figure 1 illustrates the proposed methodology.

a. Input the dataset

The customer reviews for restaurants were taken from TripAdvisor website using web crawlers. The two Restaurants each in four metropolitan cities were used as dataset in this work. Total 19,492 reviews were taken for this work.

b. Aspect extraction / selection

This provides a quick way to identify important words in the text document. The Significant words, multi-words, or regularly occurring phrases that indicate text category are commonly referred to as features or attributes. This work creates an Aspect set based on NMF topics according to the highest coherent score. The aspect set created for Restaurant reviews are Food, Service, Staff, Ambience and

Price. There are 7918, 4470,2732, 2192, and 2180 reviews respectively. The reviews of five aspects are stored in separate csv files.

c. Lexicon based Classification

Lexicon based method is used to classify the aspect-based review sentence into positive or negative. It is a rule-based opinion mining approach [7]. The aspect based collected sentences are used to classify the Sentiment using the *SentimentIntensityAnalyzer* from *nltk.vader_lexicon.* Using the method polarity_scores (), the polarity score for each sentence is classified as pos or neg. If the pos score is greater than neg, the particular sentence is added to the positive list or if vice versa, then to the negative list. Finally, both positive and negative list is stored in a csv file for each aspect and finally five csv files are generated as the output of this phase; based on these, the further steps are carried out.

d. Feature Transformation using CountVectorizer

The *sklearn.feature_extraction.text* library in Python provides the CountVectorizer. A text is transformed into a vector by counting the number of times each word occurs in the entire document [8]. This is useful when dealing with a large number of such texts and converting each word into a vector. Input texts are converted into matrices of token counts by CountVectorizer. The count matrix has a column for each known token and a row for each document; the value is many occurrences of a token in a document. After selecting features, the document is represented as a document vector, and the text classifier is trained using an appropriate machine learning technique. The training and testing data are being separated using *sklearn.model_selection - train_test_split.* 80% of the text data are for training and the remaining 20% is chosen randomly for testing purpose.

e. Single base Classifiers

1) Naive Bayes (NB)

This is an efficient classification algorithm that classifies data based on probability. These algorithms function fantastically with millions of records as well. It is based on Bayes theory and classifies data using various probabilities [9]. The predicted class in Nave Bayes is the one with the highest probability and it also referred to as Maximum a posterior. Across numerous domains, this method has various benefits and drawbacks. The approach is highly scalable and fast, and it can be used for both Multiclass and Binary Classification. It may be used on minimal datasets as well, and consequently produces good results. [10].

2) *Decision Tree*

Regression and classification can both be carried out with this algorithm. In this method, the dataset is divided in to smaller subgroups and the tree associated with each subset is generated incrementally. It is capable of handling categorical and numerical data. By using the Gini index and information gain parameter, we can choose which attribute will be used to further dataset subdivision. When the Gini index is used, the decision tree is referred to as classification and regression tree (CART), and when the information gain is used, it is referred to as ID3. A variety of applications can be handled by the algorithm [11].

3) *Support Vector Machine*

The algorithm works well for both regression and classification. To distinguish classes, it constructs a hyperplane. This approach performs well with regression, and the efficacy of SVM improves as the dimensional space increases. When the dimension number is more than the sample number, SVM performs effectively [12], however, it does not scale well to large datasets. To increase its computational efficiency, SVM extensively uses cross-validation.

For each training set, three classification algorithms are applied. The algorithms used are – Naive Bayes - *sklearn.naive_bayes.MultinomialNB,* Decision Tree - *sklearn.tree.DecisionTreeClassifier* and SVM - *sklearn.svm.LinearSVC.*For each of these algorithms, the model is created using the MultinomialNB(), DecisionTreeClassifier() and LinearSVC() respectively. The model is fit using *fit (X_train_dtm, y_train).* Then the x_test is predicted using the *predict()* function. Finally, the confusion matrix is generated and accuracy score is calculated for the independent data y using the y_test and y_predict.

f. Ensemble Classification

Ensemble methods integrate the predictions of a number of predictors to reach at a final decision. [13]. They've been split into two groups.

- Homogeneous ensemble methods
- Heterogeneous ensemble methods

1. Homogeneous Ensemble

By using adaptive (Boosting) or random (Bagging (Bootstrap and Aggregating) techniques, the homogeneous ensemble approach combines the outcomes of the same algorithm on the training set. It will improve the accuracy of the combined model after aggregating results from each model. A large number of datasets can be used for this type of ensemble method. In the homogeneous method, feature selection for different training data is the same. It is computationally expensive. In this research work, Bagging with SVM is used for homogeneous Ensemble classification.

Bagging

This ensemble approach combines the Bootstrapping and Aggregation machine learning methods into a single ensemble model. The model's high variance is reduced through the bagging method. The SVM has variance and low bias. It is sub-sampled (10 sub-samples each contains 100 samples) from a large dataset (1000 samples). On each subsample training data, multiple SVMs are constructed. While banging the sub-sampled data on the SVM, the problem of over-fitting of training data on each SVM is minimized. Each SVM is grown deep with sub-sampled training data to improve the model's efficiency. To interpret the final prediction, and the results of each SVM are combined. The variance of aggregated data starts to decline. The number of SVM used in the bagging approach affects the accuracy of the model's prediction. The different sub-samples of sample data are picked randomly with replacement.

The package used for Bagging is *sklearn.ensemble.BaggingClassifier. LinearSVC()* is used as the estimator for the Bagging Classifier because it gives the maximum accuracy compared with other Single Base Machine Learning Algorithms – Naïve Bayes and Decision Tree. The model is fit using *fit (X_train_dtm, y_train).* Then the x_test is predicted using the *predict()* function. Finally, the confusion matrix is generated and accuracy score is calculated for the independent data y using the y_test and y_predict.

2. Heterogeneous Classification

Heterogeneous classification is the combination of multiple classifiers or machine learning models, all of which are based on the same data. This strategy works well for smaller datasets. In this technique, the procedure for selecting features differs for the same training data. Using the ensemble method, the overall outcome is calculated by averaging the results of each integrated model. This research work use Stacking for heterogeneous classification.

Stacking

This strategy also integrates various classification and regression algorithms with the help of a meta-classifier or meta-model. The whole training dataset is used to train the lower-level models, and the integrated model is then trained using the results of the lower-level models. Here Parallel training is conducted for each lower-level model. The training dataset is formed by using predictions from the lower-level models as input to the next model, generating a stack in which the top layer is more trained than the bottom layer. The top-layered model, which is created from lower-level models, has a high prediction accuracy. The stack grows until the best prediction is made with the lowest error. The combined or Meta models predictions are dependence on the various weak models or lower layer model's predictions. Its goal is to build a clasiffer, which is less bias.

The package used for Stacking is *sklearn.ensemble.StackingClassifier*. The layer one estimators are the three Single Base ML classifiers (Naïve Bayes, Decision Tree and SVM). *LinearSVC()* is used as the meta classifier for the Stacking Classifier because it gives the maximum accuracy compared with other Single Base Machine Learning Algorithms – Naïve Bayes and Decision Tree. The model is fit using *fit (X_train_dtm, y_train).* Then, the x_test is predicted using the *predict()* function. Finally, the confusion matrix is generated and accuracy score is calculated for the independent data y using the y_test and y_predict.

4. Performance Metrics

The classification accuracy is the most significant measure for analysing the performance of the classifier. It's the percentile of test samples that are successfully classified. Classifier accuracy is the ability to predict the label of newly identified or previously unknown data. In the same way, the accuracy of a predictor relates to how well it can predict the value of a predicted attribute for new or previously unseen data. Accuracy detection is a key metric for measuring the classifier's efficiency. The following formula can be used to determine a classifier's accuracy

Accuracy = Number of Correctly classified records/total number of data records

The package used to calculate accuracy is from *sklearn.metrics.accuracy_score.*

5. Experimental Results

A. Dataset Description

This study makes use of a restaurant review dataset that was scraped from TripAdvisor.com. It is the largest travel community review site on the web. Retrieval of customer review data is done by crawling by using the tool beautifulsoup (vers.4).

B. Results and Discussion

Following table 1 shows the overall result of this research work. When we compare the Accuracy Percentage, in the single base classifier SVM algorithm gives high accuracy compared with the other two. In the homogenous classifier Bagging - SVM gives higher accuracy than Single Base Algorithms. Comparing all, heterogeneous algorithm – Stacking gives the highest accuracy. Following Table 1 shows the accuracy of various algorithms.

Table 1 – Algorithms' Classification Accuracy for Restaurant Reviews

Aspects	**Reviews**	**Vader+NB (%)**	**Vader+DT (%)**	**Vader+SVM (%)**	**Bagging (%)**	**Stacking (%)**
Food	7918	92.9	95.1	95.3	95.5	96.2
Service	4470	94.5	96.1	96.3	96.9	97.4
Staff	2732	97	97.3	97.4	97.6	98.2

Ambience	2180	94.4	94.6	94.9	95.1	96.4
Price	2192	85.6	90.2	90.9	91.2	91.8

The following figures 1, 2, 3, 4, and 5 show the Accuracies of different Aspects – food, Service, Staff, Price and Ambience respectively.

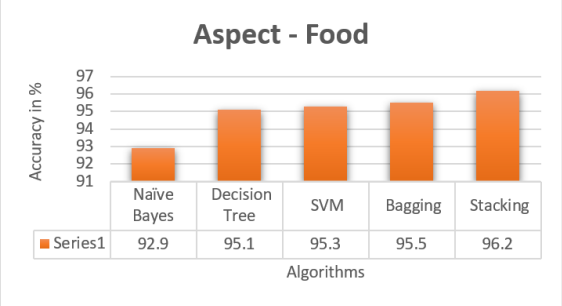

Figure 1 – Accuracy for the Aspect Food

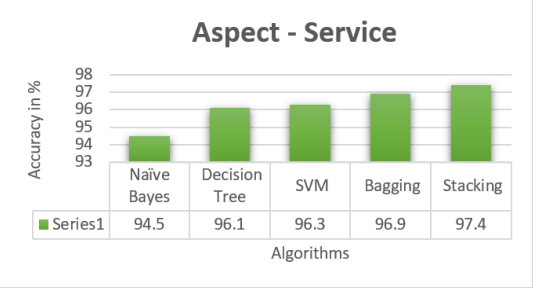

Figure 2 – Accuracy for the Aspect Service

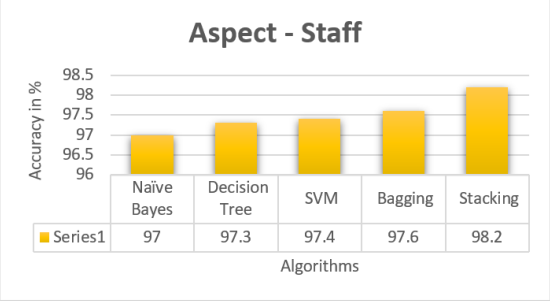

Figure 3 – Accuracy for the Aspect Staff

Figure 4 – Accuracy for the Aspect Price

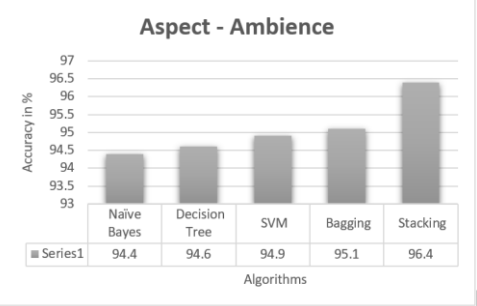

Figure 5 – Accuracy for the Aspect Ambience

We observe that the accuracy of all aspects of the heterogeneous classifier Stacking gives the maximum. In addition to that the accuracy ranges for the lowest to the highest – Single Base, Homogeneous and Heterogeneous.

Conclusion

In today's digital world, analyzing the digital data to extract user opinions is essential for increasing customer satisfaction. This work propose a set of models to evaluate these customer reviews and generate relevant feature-based opinions. To increase the performance accuracy of individual machine learning methods, this work presents ensemble techniques based on the lexicon method and machine

learning classifier. This research study evaluates individual machine learning methods Naïve Bayes, Decision Tree and Support Vector Machine, after that it performs homogenous ensemble classifier Bagging - SVM and heterogeneous ensemble classifier Stacking, on the restaurant review dataset. The single base classifier SVM gives high accuracy compared with the other two machine learning methods. The proposed homogenous ensemble classifier Bagging - SVM shown higher accuracy than Single Base classifiers. Comparing all, it is evident that the heterogeneous ensemble algorithm Stacking outperforms the others. The focus of future study will be on producing reliable opinion summaries for each aspect.

References

[1] Dietterich TG.(2000). Ensemble methods in machine learning. In: Proceedings of Multiple Classifier System. vol. 1857. Springer;(2000). pp. 1–15.

[2] M. Govindarajan, Ensemble Methods for Classification of Direct Marketing, International Journal of Management, Technology And Engineering, ISSN NO : 2249-7455, Volume 8, Issue IX, SEPTEMBER/2018

[3] Oscar Araque, Ignacio Corcuera-Platas, J. Fernando Sánchez-Rada, Carlos A.Iglesias,. Enhancing deep learning sentiment analysis with ensemble techniques in social applications, Expert Systems with Applications Volume 77, 1 July 2017, Pages 236-246

[4] Nisha Yadav, Rajeev Kumar, Dr. Bhupesh Gour, Dr. Asif Ullah Khan, Extraction-Based Text Summarization and Sentiment Analysis of Online Reviews Using Hybrid Classification Method, 978-1-7281-0012-8/19/$31.00 ©2019 IEEE

[5] Sanjeev Kumar, Dr. Ravendra Singh,. Comparative analysis of ensemble classifiers for sentiment analysis and opinion mining, 3rd International Conference on Advances in Computing,Communication & Automation, **ISBN:**978-1-5090-6403-8,2017 IEEE

[6] Azadeh Mohammadia , Anis Shaverizade,. Ensemble Deep Learning for Aspect-based Sentiment Analysis, Int. J. Nonlinear Anal. Appl. Volume 12, Special Issue, Winter and Spring 2021, 29-38 ISSN: 2008-6822 (electronic) http://dx.doi.org/10.22075/IJNAA.2021.4769

[7] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," Proc. The Eighth International AAAI Conference on Weblogs and Social media, May. 2014, pp. 2-10.

[8] Monisha V. N, Madhumitha N, Nimishamba B, Poornima N and Dr. Shashikumar D. R,. Sentimental Analysis of Movie Reviews using Twitter, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Published by, www.ijert.org NCETESFT - 2020 Conference Proceedings

[9] Ravinder Ahuja, Aakarsha Chug, Shruti Kohli, Shaurya Gupta, Pratyush Ahuja. "The Impact of Features Extraction on the Sentiment Analysis" , International Conference on Pervasive Computing Advances and Applications, Procedia Computer Science, 2019, 341–348

[10] M. Rathi, A. Malik, D. Varshney, R. Sharma, and S. Mendiratta, "Sentiment Analysis of Tweets Using Machine Learning Approach," 2018 Eleventh International Conference on Contemporary Computing (IC3), Noida, 2018, pp. 1-3

[11] YAcharya, U. R., Molinari, F., Sree, S. V., Chattopadhyay, S., Ng, K. H., and Suri, J. S., Automated diagnosis of epileptic EEG using entropies. Biomed. Signal Process. Control 7(4):401–408, 2012.

[12] İ. İşeri, Ö. F. Atasoy and H. Alçiçek, "Sentiment classification of social media data for telecommunication companies in Turkey," *2017 International Conference on Computer Science and Engineering (UBMK)*, Antalya, 2017, pp. 1015-1019

[13] Khalida DOUIBI, Nesma SETTOUTI and Mohammed El Amine CHIKH,. The Homogeneous Ensemble Methods for MLknn Algorithm, Proceedings of the International Conference on Compute and Data Analysis ,May 2017 ,Pages 197–201, https://doi.org/10.1145/3093241.3093262